

CLAIMS LISTING

Please use the following claims to replace all prior versions of the claims. Underlined text denotes added text while with a strikethrough denotes deleted text.

- 5 1. (Currently Amended) An apparatus for automatically extracting
metadata from electronic documents comprising a first processing
element, a second processing element, a reasoning element, and a
database, wherein:
- 10 i) said first processing element is configured to convert electronic
documents into PostScript files;
- ii) said first processing element is configured to provide the
PostScript files to a second processing element;
- 15 iii) said second processing element is configured to receive said
PostScript files and extract predetermined information from the PostScript
files;
- iv) said second processing element is further configured to provide
said extracted predetermined information to said reasoning element;
- v) said database is configured to provide input to said reasoning
element;
- 20 vi) said reasoning element is configured to employ a set of rules to
automatically extract metadata from the PostScript files by employing the
extracted predetermined information-and the input from the database; and
- vii) said reasoning element provides an output of metadata.
- 25 2. (Cancelled)
- 30 3. (Currently Amended) The apparatus for automatically extracting
metadata from electronic documents as set forth in claim 1, wherein said
predetermined information is substantially spatial layout facts, the spatial
layout facts being augmented strings of text, where each spatial layout fact
contains a string of text and spatial data regarding the string of text.

4. (Previously Presented) The apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein the second processing element and said database simultaneously input to the reasoning element.

5. (Previously Presented) The apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said set of rules is updated.

6. (Previously Presented) The apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said metadata is substantially comprised of title, author, affiliation, author affiliation, and table of contents.

7. (Previously Presented) The apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said metadata is provided to a user interface.

8. (Previously Presented) The apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said metadata is provided to a storage medium.

9. (Currently Amended) A method for automatically extracting metadata from electronic documents providing a first processing element, a second processing element, a reasoning element, and a database and comprising the steps of:

a) employing said first processing element to convert electronic documents to PostScript files;

b) further employing said first processing element to provide the PostScript files to said second processing element;

- c) employing said second processing element to receive said
PostScript files and extract predetermined information from the
PostScript files;
- d) further employing said second processing element to provide
5 extracted predetermined information to said reasoning element;
- e) employing said database to provide input to said reasoning
element;
- f) employing a set of rules in said reasoning element to
automatically extract metadata from the PostScript files by
10 employing the extracted predetermined information and the
input from the database; and
- g) providing an output of metadata from said reasoning element.

10. (Cancelled)

11. (Currently Amended) The method for automatically extracting
metadata from electronic documents as set forth in claim 9, wherein said
predetermined information is substantially spatial layout facts, the spatial
layout facts being augmented strings of text, where each spatial layout fact
20 contains a string of text and spatial data regarding the string of text.

12. (Previously Presented) The method for automatically extracting
metadata from electronic documents as set forth in claim 9, wherein the
second processing element and the database simultaneously input to the
25 reasoning element.

13. (Previously Presented) The method for automatically extracting
metadata from electronic documents as set forth in claim 9, wherein said
set of rules is updated.

14. (Previously Presented) The method for automatically extracting metadata from electronic documents as set forth in claim 9, wherein said metadata is substantially comprised of title, author, affiliation, author affiliation, and table of contents.

5

15. (Previously Presented) The method for automatically extracting metadata from electronic documents as set forth in claim 9, wherein said metadata is provided to a user interface.

10

16. (Previously Presented) The method for automatically extracting metadata from electronic documents as set forth in claim 9, wherein said metadata is provided to a storage medium.

15

17. (New) The apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein each string of text is bound by a bounding box and wherein the spatial data includes:

- a) a page number of the electronic document where the string of text appears;
- b) an absolute line counter order for each string of text;
- c) an x-y location of a lower left corner of a bounding box bounding the string of text;
- d) an x-y location of an upper right corner of the bounding box; and
- e) font metrics of bounding box extensions used to represent the string of text.

25

18. (New) The apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein the document includes at least a first page and text with a small font and other text with a large font, and wherein the set of rules include:

- a) a rule for extracting a title of the document, such that the title is identified as being located on an upper portion of the first page of the document and is printed using the large font;
- b) a rule for extracting authors of the document, such that authors are identified as being listed immediately under the title in some order;
- c) a rule for extracting author affiliations, such that author affiliations are identified as being located as text following the authors; and
- d) a rule for affiliating the authors with the author affiliations such that if only one affiliation appears, then all authors are associated with the one affiliation.

19. (New) The method for automatically extracting metadata from electronic documents as set forth in claim 9, wherein each string of text is bound by a bounding box and wherein the spatial data includes:

- a) a page number of the electronic document where the string of text appears;
- b) an absolute line counter order for each string of text;
- c) an x-y location of a lower left corner of a bounding box bounding the string of text;
- d) an x-y location of an upper right corner of the bounding box; and
- e) font metrics of bounding box extensions used to represent the string of text.

20. (New) The method for automatically extracting metadata from electronic documents as set forth in claim 9, wherein the document includes at least a first page and text with a small font and other text with a large font, and wherein the set of rules include:

- a) a rule for extracting a title of the document, such that the title is identified as being located on an upper portion of the first page of the document and is printed using the large font;

- b) a rule for extracting authors of the document, such that authors are identified as being listed immediately under the title in some order;
- c) a rule for extracting author affiliations, such that author affiliations are identified as being located as text following the authors; and
- 5 d) a rule for affiliating the authors with the author affiliations such that if only one affiliation appears, then all authors are associated with the one affiliation.

10

15

20

25

30